DOCUMENT RESUME

ED 161 938 TM 007 973

AUTHOR Gowin, D. Bob; Millman, Jason
TITLE Can Meta-Evaluation Give a Direction for Research on Evaluation?
PUB DATE 29 Mar 78
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 29, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Documentation; *Evaluation; *Evaluation Criteria; *Evaluation Methods; Information Utilization; *Values
IDENTIFIERS *Meta Evaluation

ABSTRACT

Describing evaluation in three tiers (the events or objects that are the subject of evaluation, the reports and other documents about that subject, and the analysis of those documents), this report focuses on the third tier, meta-evaluation. The crux of evaluation is the concept of value, having five major forms of value claims: the intrinsic claim that something is good, the instrumental claim that it is useful, the comparative claim that it is better than something else, the decision claim that it should be chosen, and the idealization claim that it could be better. Each kind of claim requires a different appropriate analysis. Valuing is defined as a caring relationship between persons and objects. Evaluation criteria develop as a consequence of disciplined inquiry; the subject matter studied and the criteria used influence each other. Three procedures are suggested as aids for meta-evaluation. One is the use of standards; another is the use of a checklist. The third is the use of a set of five questions: What is the significant question of the study and its assumptions? What is it about? What methods were used? What answers were given? And what is the conceptual structure of the work? (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED161938

TM007 973

Printed in U.S.A.
Session No. 17.26

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JASON MILLMAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Can Meta-Evaluation Give a Direction for Research on Evaluation?[1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

D. Bob Gowin and Jason Millman

Cornell University

Yes, it can. We argue for four ways. By meta-evaluation we refer to the study of evaluations. Perhaps we should call our point of view evaluation appraisal, or evaluation critique. We end with a careful critique of evaluations.

We believe that meta-evaluation should go beyond a mere recycling of the conventional wisdom of evaluation. The sources for the critique are not limited to what a study of evaluation simpliciter would produce. We bring into the critique of evaluation ideas from value theory, from epistemology, from ethical and social philosophy, from theories of education, as well as the typical sources of research methodology and statistics, evaluation standards and practices,[2] and research on empirical summative evaluation.[3]

Our starting point is with any and all documentation of evaluation

---

[1]Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, Ontario, 29 March 1978.

[2]Daniel Stufflebeam, "Meta Evaluation." Unpublished paper, February 1978.

[3]Thomas D. Cook and Charles L. Gruder, "Metaevaluation Research," Evaluation Quarterly, Vol. 2, No. 1, February 1978, pp. 5-51.

work. We lay on the table these documents about evaluation studies. We, in turn, analyze these documents.[4]

We have here three tiers. We can become very confused if we do not maintain a firm linguistic hold on our language about these three levels. At the bottom is whatever is being evaluated. It could be a program, a product, a process, whatever. If a science program in middle schools is being evaluated, then that science program is the first layer. The second layer consists of all the plans, reports, records, interview data, and other documentation generated by the activities of the evaluation study. These documents are the second tier. Next, those materials are analyzed. In doing this analysis we are doing meta-evaluation. The report of that analysis is the third tier. So we have three tiers: first, the event or objects (e.g., the science program); second, the evaluation of that event or object (e.g., the evaluation of the program); third, the analysis of the evaluation.

This paper focuses on the third tier. Meta-evaluation, the third tier, is supposed to produce added insights with respect to the evaluation documentation, the second tier. Stufflebeam distinguishes between formative meta-evaluation which is done "to insure that the primary evaluation will be done as adequately, efficiently, fairly, and usefully as possible" and summative meta-evaluation designed "to publicly expose strengths and weaknesses of completed evaluation and thereby to protect practitioners and consumers from the application of faulty conclusions and recommendations."[5]

---

[4]"Document" is one word that refers to products, reports and recorded activities of evaluators. We include here for convenience those informal and often unrecorded discussions with evaluators, workshops on evaluation, and other activities that lead to good ideas about evaluation.

[5]Stufflebeam, op. cit., p. 6.

We wish to argue here that in addition to such anticipated benefits of meta-evaluation, there are other benefits that come from analyzing evaluation documents that transcend the particular evaluation context under scrutiny. We believe that meta-evaluation can help us get smart about doing evaluations by contributing to four areas of research about evaluation. From meta-evaluations, from analyzing evaluations, we can learn more about: (1) the practice of evaluation, (2) the concept of value, (3) the criteria for judging evaluations, and (4) the procedures for conducting meta-evaluations. Each of these directions for research will be discussed in turn.

Part I: The Practice of Evaluation

Perhaps the most obvious contribution meta-evaluation offers to research in evaluation is to make us better informed about the state of the art of evaluation work. Data-based descriptions and statements of relationships can be sought. Previous papers in this symposium have suggested some directions for such research. Among other tasks, we can: (a) identify instances in which new methods are applied to evaluative tasks; (b) investigate the circumstances under which procedurally sound evaluations are most likely to take place; (c) determine the correlates of evaluations that seem to have impact; and (d) find out who evaluates for whom and how and why the evaluations are performed.

We describe the practice and also explain why the practice does or does not make sense by citing the reasons given to justify it. But the ultimate goal in studying the practice of evaluation is to understand it and improve it in much the same way that the goal of many evaluations

(the second tier) is to understand and improve programs or products (the first tier).

Part II: The Concept of Value

Evaluation is often defined as the assessment of worth. Worth, however, is just as often undefined. Worth is another word for value. So it should be clear that evaluation is the assessment of value, and that the concept of value is a key concept in evaluation work. Unfortunately, value is a much-neglected aspect of evaluation. This fact becomes very clear when we undertake meta-evaluation analysis. We next will present some of the basic problems about value in evaluation, present ideas from value theory, and conclude with a few things we have found in the past year of our studies. A task in meta-evaluation is to clarify concepts of value in order to develop sharper, more telling questions in evaluation work.

Epistemology of evaluation.

Knowledge claims and value claims ride in the same boat. Both kinds of claims establish the relation between concepts, events-objects, and the records of those events we call facts. For a simple but significant epistemological heuristic, think of a vee. At the bottom point of the vee is the event (or object) we want to get smart about. Up the left arm of the vee is the key concept which signifies regularities in the event. Up the right arm of the vee is the record we make of the event or the observations we make of the object, i.e., the facts.[6] For example, a chemist with the

---

[6] D.B. Gowin, Chapter 4 in The Domain of Education. Unpublished manuscript, 1977.

concept of "boiling point" can make an event of boiling of a liquid happen, and make a record of that event by recording thermometer readings. The concept is "boiling point," the event is the boiling of the chemical, and the facts are the thermometer records.

As their chief product, scientific research produces knowledge claims, and evaluation produces value claims. But on the way to producing their chief products, science produces value claims, and evaluation produces knowledge claims. Confusing, but true. Scientific research is shot through with evaluation. The concepts, method, and techniques of research are interlaced with value judgments. Acts of evaluation may be among the most significant acts in scientific work.[7] Evaluation studies and evaluation research cannot be bereft of all values, value judgments, and value concepts; evaluation is not value-free. Value theory, when applied to evaluation practice, should show us what our values are.

Value theory.

Philosophers often categorize their work as theories of knowledge (epistemology), theories of value (axiology), and theories of reality (metaphysics). A theory of knowledge is relevant to how we make knowledge claims, and a theory of value is relevant to how we make value claims.

In value theory there are a number of elementary distinctions relevant to evaluation. Value theory, like any good theory, gives us a set of conceptual goggles which provide a comprehensive view of the phenomena of interest. Once we begin to see the place of value theory in evaluation

---

[7] M. Scriven, "Evaluation in Science Teaching," Journal of Research in Science Teaching, Vol. 13, No. 4, 1976, pp. 363-88. Scriven argues that "evaluation is an absolutely crucial part of science" although it has been "closet evaluation."

work, we begin to make connections within our domain of interest that we have not seen explicitly before, may not have been aware of before.

Since works of evaluation are virtually saturated with different kinds of value claims, we need systematically to sort them out, and value theory helps with this sorting task in identifying separate kinds of claims. In all the evaluation studies we have appraised so far, we find five major forms of value claims. These are:

1. Claims that X is good; the intrinsic value claim.
2. Claims that X is good for something else; the instrumental value claim.
3. Claims that X is better than Y; the comparative value claim.
4. Claims that X is right; claims that you ought to do, to choose or to decide; the decision claim form.
5. Claims about how X could be better; seeing in what is that which would be better; the idealization claim form.

Statistical theory gives us different analyses for different kinds of statistical tests. Value theory gives us different analyses for different kinds of value claims. The evidence and argument to support an instrumental value claim are different from the evidence and argument needed to support a decision claim. Each of the five claims requires a different backing. We can do little more here than just illustrate these distinctions. We take up three points. The first point concerns the difference between the good and the right, the second sketches a difference between intrinsic and instrumental value claims, and the third suggests a connection between the definition of valuing behavior and its place in evaluation studies.

The good and the right. Philosophers do not agree (as is the case with specialists in any field) about the good and the right. Nevertheless, we can profitably use some basic distinctions in evaluation work.

Of a science program or a washing machine we could ask a number of questions: Is X good? (intrinsic claim). What is X good for? (instrumental claim). Is X better than Y? (comparative claim). Is X good enough? Can we make X better? Is X the best sort of thing X can be? (idealization claims). These are claims about the good of the object or event we are studying in evaluation. Such claims will be found in evaluation reports. This group of claims about the good composes the largest number of value claims.

If we think evaluation studies produce a basis for a decision, then we find value propositions in evaluation studies which make a claim that X is right, that we ought to do X, or that we ought to choose X, or choose X over Y. In general these claims assert that X is the right thing to do, to choose, to decide.

For example, we may be evaluating a program of special education. Is a decision for special education right? Is it right to spend more money on handicapped than normal children? Yes, such decisions are right if we accept a value principle of justice: treat unequals unequally and equals equally.

Claims about what is right are sometimes closely related to people's rights. To claim a person has a right is to claim that s/he is entitled to behave in certain ways in certain circumstances. Typically "rights" define expectations (and responsibilities) of other people toward the possessor

of the right.[8] Evaluation studies often are attempts to educate people about their rights and their responsibilities for the rights of others (e.g., Title IX evaluations).

Evaluation research must find systematic ways of asking and answering these questions about what is taken to be the good and the right. There are many different techniques evaluators can be trained to use. One of these techniques concerns the form of the practical syllogism in which an argument can be constructed which relates a value principle to facts, and relates both of these to a conclusion, or decision about what one ought to do or choose.[9] This skill would be most useful for evaluators to develop.

Intrinsic value and instrumental value. Evaluation is the assessment of worth. Such assessments involve value in two ways. First, the assessment itself is a value judgment, a value claim. Second, what the assessment is about is a value.

A value is the good of some object or event. When we talk about the good of something we are talking about its intrinsic value. For example, the good of mathematics is found in its proofs, its methods, the internal relations of one part of math to another part. Such claims about the good of mathematics involve us in careful descriptions (the attempt to say what it is, its existence), plus the judgment that those descriptions are about the value aspects of mathematics. That is, not all descriptions of mathematics are about its values; some are about its history, for example. We do

---

[8] An elementary and clear treatment of concepts of value (including rights, utility, justice) is found in Peter A. Facione et al., Values and Society. An Introduction to Ethics and Social Philosophy. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1978.

[9] For a simple but not simple-minded discussion see Jerrold Coombs, "Objectives of Value Analysis," in L. Metcalf, Values Education, 41st yearbook, National Council for the Social Studies, 1971, pp. 1-29.

not insist that all observers agree about the good of mathematics; there can be competing claims of the good. The main point is this: assessments of the good of something are assessments of intrinsic value. This notion of intrinsic value seems to be one of the most difficult of all of the concepts about value.[10]

Perhaps the easiest of all of our concepts about value concerns instrumental value claims. An instrumental value is the good of something for something else. Mathematics has instrumental value for engineering, for physics, for medicine. Mathematics must be good in itself, have intrinsic value, for it to serve the office of being a good instrument for engineering, have instrumental value. Sometimes an instrument that is bad for some purposes (e.g., a broken hammer) can be instrumentally-good for other purposes (e.g., a paper weight). Bad mathematics may make good engineering claims. But in any case we separate through analysis just what the good (or bad) of the object is from the instrumental good the object serves. This point is related to the use of consequences as a test of goodness, the ends-means relation, and the notion of contributory good (good as an ingredient in something else).

Valuing. One of the key concepts of value theory is "valuing." Valuing is defined as a caring relation between persons and objects (events). Valuing is a process (of behavior/action) of human beings who strive to keep in existence something that exists independently of their caring. Valuing behavior is studied in some evaluations. For example, if an evaluator wants to establish credibility with clients, then finding

---

[10]Robert Hartman, The Structure of Value. S. Illinois University Press, 1967, p. 384.

out their real questions and answering those can establish credibility. The clients' "real questions" provide evidence of their valuing behavior: it shows what they really care about. When an evaluation report is significantly related to the valuing behavior of clients, the likelihood of it having an impact on the clients increases. This relation between valuing behavior and the evaluation report can be studied directly in research on evaluation.

Part III: The Creation and Development of Criteria to Use in Evaluation

One way to look at progress in science is to trace the development of criteria used to shape knowledge claims. For example, one person's observation of an event is not sufficient in many contexts, we now think, so we use criteria of intersubjective agreement to replace simple observation. But human reporters are more fallible than mechanical or electronic devices we can use to make records of events, so we work at instrument calibration. There are times, however, when the best instrument is a well-trained person, a well-calibrated human being. These various criteria are generated in the course of scientific practice. We judge events according to our criteria; and we invent criteria to judge events. This alternation between using stable criteria, and changing criteria to fit new problems, is part of scientific practice.

We borrow from scientific practice to see if we can find analogues in the practice of evaluation. Where do our criteria come from? Many of them are created on the spot; they are inventions of practical necessity. Once invented, we can test their wider applicability in subsequent studies. Once invented, criteria can be extracted from the context in which they first were used, and implanted in other contexts.

Criterial forms accrue to evaluation research as a consequence of disciplined inquiry.[11] We shape the subject matter by our criteria, and our subject matter supplies us with the shape of our criteria. This is no vicious circle, but rather an alternation of focus of inquiry. We shape our instruments, then we use them. The use of the instruments alters their shape.

It is one of the tasks of meta-evaluation to show how this process of bootstrapping works.

Part IV: The Procedures for Conducting Meta-Evaluations

The task of conducting a meta-evaluation requires procedures or tools that can be used to perform the analyses. A fertile area for research in evaluation is to develop guides to doing competent meta-evaluations.

Our own development work has proceeded along three separate tacks. In each case meta-evaluators are given a set of concerns about evaluations that focuses their attention on critical features of the evaluation documents. In one technique, the meta-evaluator uses standards for educational evaluations as the criteria against which to judge the original evaluation effort. An impressive start has already been made by a multi-association commission to identify important standards.[12] The other two guides are (1) a modification for meta-evaluation of Michael Scriven's checklist for the evaluation of products,[13] and (2) a set of questions that is designed to expose the logical structure of the original evaluation.[14]

---

[11] Dewey, Logic, Chapter 1,6.

[12] Stufflebeam, op. cit.

[13] M. Scriven, "Evaluation Perspectives and Procedures," Chapter 1 in W. James Popham (Ed.), Evaluation in Education: Current Applications. Berkeley, Calif.: McCutchan, 1974.

[14] Gowin, op. cit.

Checklist guide.

In the checklist guide, the assumption is made that the meta-evaluator should view the evaluation documents both as tier two and tier one products. Typically, the meta-evaluator focuses on the evaluation as tier two documents and considers how soundly the evaluator judged the adequacy of the original or tier one program or product. That is, usually the meta-evaluator evaluates the quality of the tier two evaluation work, and asks whether the evaluator followed approved procedures, made justifiable claims, etc.

But, evaluations consist of activities, reports, and events and can be considered tier one programs or products just like the middle school science program. As a tier one product, the evaluation is judged against the same criteria that are used with other programs and products. We ask, for example, if the evaluation was needed and if the evaluation was cost effective.

Thus, when the checklist guide is used, there are two foci or objects for the meta-evaluation; namely (1) the judgments of the tier two evaluation, and (2) the evaluation considered as a tier one product or program. When focused upon the tier two evaluation claims, the checklist addresses the quality with which the evaluator assessed such components of the tier one program as its need and market, the technical adequacy of the performance data and other indicators of program effect, and its cost and benefit. The questions asked about the evaluation when it is considered as a tier one program or product in its own right address such concerns as the need for the evaluation, the appropriateness of the agreements under which the evaluation was conducted, the intended outcomes and side

effects of the evaluation, existence of impact reducing factors, and the evaluation's cost and benefit.

QUEMAC.

The acronym QUEMAC has been created to name the approach to meta-evaluation which seeks to detect the logical structure of the evaluation.[15] QUEMAC stands for Question, Event-object, Method, Answer, and Concept. The meta-evaluator following the QUEMAC guide, answers such questions as:

1. What is the significant, telling question of the study? What are the unquestioned assumptions?
2. What is the evaluation about? What events or objects are described?
3. What methods and techniques were used to answer the key questions?
4. What answers were given? What were the major claims of the evaluation?
5. What key concepts does the evaluation use? What is the conceptual structure of the work?

By answering these five questions about any evaluation study we produce a brief critical abstract of the study. The QUEMAC appraisal reveals the logic which connects the separate parts of the evaluation study. It is a reconstructed logic to be sure. That is, it is the logic of the appraisal and not of the makers of the original evaluation study.

The simplest logic connects the question asked with the answer given. QUEMAC appraisals led us to one interesting discovery. The key questions in evaluation studies are questions which have value concepts embedded in them. For example, we might ask: "Is achievement in this reading program better than

---

[15] Gowin, op. cit.

the national average?" So easy is it to suppose that more of something (higher numbers) also means better than, that we often fail to see the value concepts embedded in evaluation questions.

Learning to become explicit about these concepts of value in our evaluation questions should help both in the planning of evaluations (formative) and the reporting of completed studies (summative).

Summary.

We said in the beginning that our point of view about meta-evaluation concerned evaluation critique,[16] or evaluation appraisal. We wrote about the description, explanation and improvement of the practice of evaluation. We spent some pages to show value theory can help in evaluation appraisal and, when applied critically to evaluation practice, should show us how we treat value questions and also to exhibit what our values are. We believe that evaluation research should find systematic ways of asking and answering these questions about the good and the right because we believe we will be asking more telling questions in evaluation studies. We suggested that the notion of valuing, as it comes to us from value theory, is an important idea for evaluation practice. We wrote about bootstrapping in the invention and use of criteria in evaluation.

Finally, we presented briefly three approaches to meta-evaluation. One concerns the standards for evaluation. The second is a checklist of questions for evaluation. The third, named QUEMAC, is a series of questions which, when answered, reveals the logical structure of evaluation studies.

---

[16] For an excellent example of meta-evaluation critique see Ernest R. House et al., "Critiquing a Follow Through Evaluation," Phi Delta Kappan, March 1978, pp. 473-4.